

#4

520.38682X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE



Applicant(s): Shinichi NAKAYAMA, ET AL.  
Serial No.:  
Filed: June 30, 2000  
Title: STORAGE SUBSYSTEM AND STORAGE CONTROLLER  
Group:

LETTER CLAIMING RIGHT OF PRIORITY

Honorable Commissioner of  
Patents and Trademarks  
Washington, D.C. 20231

June 30, 2000

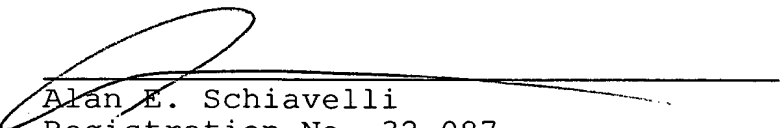
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Patent Application No.(s) 11-353806 filed December 14, 1999.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

  
Alan E. Schiavelli  
Registration No. 32,087

AES/rdh  
Attachment  
(703) 312-6600

日本国特許庁

PATENT OFFICE  
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application:

1999年12月14日

出願番号  
Application Number:

平成11年特許願第353806号

出願人  
Applicant(s):

株式会社日立製作所

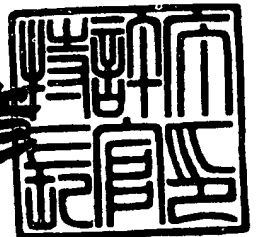


CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 5月19日

特許庁長官  
Commissioner,  
Patent Office

近藤隆彦



【書類名】 特許願

【整理番号】 PNT991022

【提出日】 平成11年12月14日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージシステム事業部内

【氏名】 中山 信一

【発明者】

【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージシステム事業部内

【氏名】 横畑 静生

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100068504

【弁理士】

【氏名又は名称】 小川 勝男

【電話番号】 03-3661-0071

【選任した代理人】

【識別番号】 100086656

【弁理士】

【氏名又は名称】 田中 恭助

【電話番号】 03-3661-0071

【手数料の表示】

【予納台帳番号】 081423

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶サブシステム及び記憶制御装置

【特許請求の範囲】

【請求項 1】

上位外部から受け取った入出力要求に応じて下位外部との間で入出力データの転送を制御する記憶制御装置において、前記記憶制御装置は、上位外部とのインタフェースに応じて前記入出力要求を受け取る少なくとも 1 台の外部インタフェースコントローラと、前記入出力要求の処理をする少なくとも 1 台の制御プロセッサと、前記外部インタフェースコントローラと前記制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループとを有することを特徴とする記憶制御装置。

【請求項 2】

前記外部インタフェースコントローラの上位外部とのインタフェースは、ファイバチャネルインタフェースであることを特徴とする請求項 1 記載の記憶制御装置。

【請求項 3】

前記外部インタフェースコントローラは、ファイバチャネルインタフェースと異なる上位外部とのインタフェースと前記ループのファイバチャネルインタフェースとの間のインタフェース変換機能を有することを特徴とする請求項 1 記載の記憶制御装置。

【請求項 4】

前記ループは、入力信号に応じて前記外部インタフェースコントローラと前記制御プロセッサとの間の伝送経路を形成する電子スイッチ機構を有することを特徴とする請求項 1 記載の記憶制御装置。

【請求項 5】

上位外部から受け取った入出力要求に応じて下位側の記憶媒体の駆動装置との間で入出力データの転送を制御する記憶サブシステムにおいて、前記記憶サブシステムは、上位外部とのインタフェースに応じて前記入出力要求を受け取る少なくとも 1 台の外部インタフェースコントローラと、前記入出力要求の処理をする

少なくとも 1 台の制御プロセッサと、前記外部インタフェースコントローラと前記制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループとを有することを特徴とする記憶サブシステム。

【請求項 6】

上位外部から受け取った入出力要求に応じて下位外部との間で入出力データの転送を制御する記憶制御装置において、前記記憶制御装置は、前記入出力要求を受け取る複数の外部インタフェースコントローラと、前記入出力要求の処理をする複数の制御プロセッサと、前記外部インタフェースコントローラと前記制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループとを有し、前記制御プロセッサは、前記ループを介して送られる前記入出力要求のうち自プロセッサ宛てのアドレスを有するフレームを読み込む手段と、読み込んだフレームについて前記入出力要求の処理をする手段とを有することを特徴とする記憶制御装置。

【請求項 7】

上位外部から受け取った入出力要求に応じて下位外部との間で入出力データの転送を制御する記憶制御装置において、前記記憶制御装置は、前記入出力要求を受け取る複数の外部インタフェースコントローラと、前記入出力要求の処理をする複数の制御プロセッサと、前記外部インタフェースコントローラと前記制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループと、前記制御プロセッサによって共通にアクセスされ前記制御プロセッサ対応に処理すべき入出力要求の論理ユニット番号を記憶する記憶手段とを有し、前記制御プロセッサは、他の制御プロセッサの稼動状態を監視する手段と、他の制御プロセッサが停止していることを検出したとき、前記記憶手段上の前記論理ユニット番号を書き換えることにより、停止した制御プロセッサに関する処理を引き継ぐ手段とを有することを特徴とする記憶制御装置。

【請求項 8】

前記記憶手段は、前記制御プロセッサ対応に前記ループ上の物理アドレスと処理すべき入出力要求の論理ユニット番号とを記憶し、前記制御プロセッサは、他の制御プロセッサが停止していることを検出したとき、前記記憶手段上の前記物

理アドレスと前記論理ユニット番号とを引き継ぐべく書き換えることにより、停止した制御プロセッサに関する処理を引き継ぐ手段とを有することを特徴とする請求項 7 記載の記憶制御装置。

【請求項 9】

上位外部から受け取った入出力要求に応じて下位外部との間で入出力データの転送を制御する記憶制御装置において、前記記憶制御装置は、前記入出力要求を受け取る複数の外部インタフェースコントローラと、前記入出力要求の処理をする複数の制御プロセッサと、前記外部インタフェースコントローラと前記制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループと、前記制御プロセッサによって共通にアクセスされ前記制御プロセッサ対応に処理すべき入出力要求の論理ユニット番号を記憶する記憶手段とを有し、前記制御プロセッサは、処理入出力要求数を計数する手段と、他の制御プロセッサに処理入出力要求数を通知する手段と、他の制御プロセッサの処理入出力要求数を取得する手段と、制御プロセッサ間の処理入出力要求数が平均化するように前記記憶手段上の前記論理ユニット番号を書き換える手段とを有することを特徴とする記憶制御装置。

【請求項 10】

上位のホストコンピュータから受け取った入出力要求に応じて下位側の記憶媒体の駆動装置との間で入出力データの転送を制御する記憶サブシステムにおいて、前記記憶サブシステムは、上位ホストコンピュータとのインタフェースに応じて前記入出力要求を受け取る少なくとも 1 台の外部インタフェースコントローラと、データを一時的に格納するキャッシュメモリと、前記入出力要求を解析し、前記ホストコンピュータと前記キャッシュメモリとの間で行う入出力データの転送を制御する少なくとも 1 台の上位側の制御プロセッサと、前記外部インタフェースコントローラと前記上位側の制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループと、前記キャッシュメモリと前記記憶媒体の駆動装置との間で行う入出力データの転送を制御する少なくとも 1 台の下位側の制御プロセッサと、前記下位側の制御プロセッサと前記記憶媒体の駆動装置との間に介在し、下位駆動装置とのインタフェースに応じ

て前記記憶媒体の駆動装置との間で入出力データの転送を行うドライブインタフェースコントローラとを有することを特徴とする記憶サブシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、上位側でホストコンピュータと接続する記憶サブシステム及び記憶制御装置に係わり、特に処理性能及び信頼性の向上を図る記憶サブシステム及び記憶制御装置に関する。

【0002】

【従来の技術】

近年、コンピュータシステムの大規模化、データ処理の高速化、24時間あるいは365日無停止運転の必要性、データ転送インタフェースの高速化などに伴い、記憶制御装置に対して性能、信頼性、可用性の向上が強く要求されている。性能向上を目的とする記憶制御装置の例として、記憶制御装置内に内部ネットワークを備えることにより性能向上を図ることを目的とする特開平11-7359号公報に開示される技術がある。

【0003】

また記憶制御装置に接続するホストコンピュータとの間のインタフェースについて着眼すると、図8に示すように複数種類のインタフェースをもつホストコンピュータを接続する必要があるが生じている。記憶制御装置内のホストインタフェース部は、ホストインタフェースごとに設けられ、その制御プロセッサは、ホストコンピュータから受け取った入出力コマンドを解析し、信号線を介してキャッシュメモリ215との間でデータ転送を行う。例えば特開平9-325905号公報はこのような記憶制御装置を開示する。

【0004】

さらに近年、性能及び可用性を向上させるために、ホストコンピュータと記憶制御装置との間のインタフェースをSCSI (Small Computer System Interface) からファイバチャネルインタフェースに置き換えた公知技術が知られている。例えば特開平10-333839号公報は、記憶制御装置とホストコンピュータ



間をファイバチャネルインタフェースによって接続する技術を開示する。これはファイバチャネルインタフェースを有するホストコンピュータ専用の記憶制御装置である。

【 0 0 0 5 】

【発明が解決しようとする課題】

上記の特開平 1 1 - 7 3 5 9 号公報および特開平 9 - 3 2 5 9 0 5 号公報に開示される技術は、ホストコンピュータから受け取った入出力要求を一つの制御プロセッサが処理するため、制御プロセッサの性能によって記憶制御装置全体の性能が押さえられるという問題がある。また制御プロセッサの障害によって関連するホストコンピュータから記憶制御装置が使用できなくなるという問題がある。特に昨今のファイバチャネルは 1 0 0 M B / S という高速データ転送が可能であるため、制御プロセッサの処理性能がネックとなってファイバチャネルのもつデータ転送速度を充分生かせない。

【 0 0 0 6 】

さらに特開平 1 0 - 3 3 3 8 3 9 号公報に開示された技術はファイバチャネルインタフェース専用の記憶制御装置であるため、S C S I インタフェースを有するホストコンピュータを接続することができない。

【 0 0 0 7 】

本発明の目的は、記憶制御装置の性能向上、特にファイバチャネルのもつ高速データ転送を生かすような高性能をもつとともに、信頼性、可用性の高い記憶サブシステム及び記憶制御装置を提供することにある。

【 0 0 0 8 】

本発明の他の目的は、複数種類のインタフェースをもつホストコンピュータを接続可能な記憶サブシステム及び記憶制御装置を提供することにある。

【 0 0 0 9 】

【課題を解決するための手段】

本発明は、上位外部から受け取った入出力要求に応じて下位側の記憶媒体の駆動装置との間で入出力データの転送を制御する記憶サブシステムにおいて、この記憶サブシステムは、上位外部とのインタフェースに応じて入出力要求を受け取

る少なくとも 1 台の外部インタフェースコントローラと、入出力要求の処理をする少なくとも 1 台の制御プロセッサと、外部インタフェースコントローラと制御プロセッサとの間に介在し、両者間の情報伝送の伝送路となるファイバチャネルインタフェースのループとを有する記憶サブシステムを特徴とする。またこのような記憶制御装置を特徴とする。

#### 【0010】

また本発明は、上位外部とのインタフェースがファイバチャネルインタフェースであるような外部インタフェースコントローラ、およびファイバチャネルインタフェースと異なる上位外部とのインタフェースとファイバチャネルインタフェースとの間のインタフェース変換機能を有する外部インタフェースコントローラを設けた記憶サブシステム及び記憶制御装置を特徴とする。

#### 【0011】

#### 【発明の実施の形態】

以下、本発明の実施形態について図面により詳細に説明する。

#### 【0012】

図 1 は、本発明の一実施形態を示すディスクサブシステムを含むシステムの構成図である。ディスク制御装置 107 は、上位外部側で、ホストコンピュータ 100、101 及び 102 と接続する。ホストコンピュータ 101 は、メインフレーム系のコンピュータであり、メインフレーム系チャネルを介してディスク制御装置 107 と接続する。ホストコンピュータ 100 は、オープンシステム系のホストコンピュータであり、ファイバチャネルインタフェースを介してディスク制御装置 107 と接続する。ホストコンピュータ 102 は、オープンシステム系のコンピュータであり、SCSI (Small Computer System Interface) を介してディスク制御装置 107 と接続する。ディスク制御装置 107 は、下位外部側でファイバチャネルインタフェースのループ 125 及び 126 を介してドライブ 127、128、129 及び 130 と接続する。

#### 【0013】

ホストインタフェースコントローラ (HIFC) 103、104 及び 105 は、それぞれホストコンピュータ 100、101 及び 102 と接続し、また他方で

ファイバチャネルインタフェースに準拠するループ 133 と接続する。制御プロセッサ 114, 115, 116 及び 117 は、一方でループ 133 と接続し、また他方で共通バス 118 と接続する。共通バス 118 には、制御プロセッサ 114 ~ 117 のほかに共通制御メモリ 112、キャッシュメモリ 122、制御プロセッサ 119 及び 120 が接続される。制御プロセッサ 119 および 120 は、それぞれファイバチャネル 141 を介してドライブインタフェースコントローラ (DIFC) 123 及び 124 と接続する。DIFC 123 および 124 は、それぞれループ 125 及び 126 を介してドライブ 127, 128, 129 及び 130 と接続する。制御プロセッサ 114, 115, 116 及び 117 は、信号線 132 を介してサービスプロセッサ 131 と接続する。

#### 【0014】

HIFC 103 は、上位外部との間のインタフェースコントローラであり、ホストコンピュータ 100 からフレームの形式で受け取った入出力コマンド、データ及び制御情報をそのままの形式でループ 133 を介して制御プロセッサ 114 ~ 117 のいずれかへ転送する。またループ 133 を介して制御プロセッサ 114 ~ 117 からフレームの形式で受け取ったデータ及び制御情報をそのままホストコンピュータ 100 へ転送する。HIFC 104 は、ホストコンピュータ 101 から受け取ったチャネルコマンド、データ及び制御情報をファイバチャネルのフレーム形式に変換し、ループ 133 を介して制御プロセッサ 114 ~ 117 のいずれかへ転送する。また制御プロセッサ 114 ~ 117 からフレームの形式で受け取ったデータ及び制御情報をメインフレーム系のチャネルインタフェースに従うデータ形式に変換してホストコンピュータ 101 へ転送する。HIFC 105 は、ホストコンピュータ 102 から受け取った入出力コマンド、データ及び制御情報をファイバチャネルのフレーム形式に変換し、ループ 133 を介して制御プロセッサ 114 ~ 117 のいずれかへ転送する。また制御プロセッサ 114 ~ 117 からフレームの形式で受け取ったデータ及び制御情報を SCSI に従うデータ形式に変換してホストコンピュータ 102 へ転送する。1 台の HIFC 103, 104 又は 105 にそれぞれ複数台のホストコンピュータ 100, 101 及び 102 を接続することも可能である。

## 【0015】

キャッシュメモリ 122 は、共通バス 118 のバスインタフェースを介してすべての制御プロセッサ 114～117, 119 及び 120 からアクセス可能なメモリであり、ホストコンピュータ 100～102 から送られたデータ及びドライブ 127～130 から読み出したデータを一時的に格納するために用いられる。キャッシュメモリ 122 上のデータは、キャッシュスロットと呼ばれるデータ管理単位に分割されている。

## 【0016】

共通制御メモリ 112 は、共通バス 118 を介してすべての制御プロセッサ 114～117, 119 及び 120 からアクセス可能な共通メモリであり、制御プロセッサ間の通信のための領域、キャッシュスロット管理テーブルなどのほかに、制御プロセッサ 114～117 の各々がループ 133 を介して取り込むべきフレームを設定する FCAL (Fibre Channel Arbitrated Loop) 管理情報 113 を格納する。

## 【0017】

制御プロセッサ 114～117 の各々は、共通制御メモリ 112 上の FCAL 管理情報 113 を参照し、ループ 133 上を流れるフレームのうち設定されたアドレスをもつフレームを取り込み、入出力コマンドによって指定された入出力要求を実行する。すなわちリードコマンドの場合には、キャッシュメモリ 122 上に要求されたデータがあれば読み出して、ループ 133 及び HIFC 103～105 のいずれかを介して要求元のホストコンピュータへ送信する。キャッシュメモリ 122 上に要求されたデータがなければ、制御プロセッサ 119 及び 120 に対して入出力要求を通知する。またライトコマンドの場合には、書き込みデータをキャッシュメモリ 122 上のキャッシュスロットに書き込み、制御プロセッサ 119 及び 120 に対して入出力要求を通知する。

## 【0018】

制御プロセッサ 119 及び 120 は、制御プロセッサ 114～117 から入出力要求の通知を受け取り、リードコマンドの場合にはドライブ 127～130 から要求されたデータを読み出してキャッシュメモリ 122 上のキャッシュスロ

トに書き込む。またライトコマンドの場合には、キャッシュメモリ 122 上のデータをドライブ 127～130 上に書き込む。

#### 【0019】

図2は、HIFC103～106と制御プロセッサ114～117との間に介在し、両者間の伝送路となるループ133及び関連機構の構成を示す図である。ループ133は、PBC (Port Bypass Circuit) 108, 109, 110及び111を有し、いわゆるハブ (HUB) 構造を形成している。PBC108～111は、各々1入力n出力の電子スイッチであり、図示するようにHIFC103～106及び制御プロセッサ114～117と接続するとともに、PBC108、PBC111間及びPBC109、PBC110間を接続する。本例のPBC108～111は、1入力2出力のスイッチであり、PBCに入力信号を与えると出力経路を限定することができる。ファイバコントローラ (FC) 151は、制御プロセッサ114～117の各々に前置するファイバチャネルインタフェースコントローラであり、ループ133を介して送られるフレームの宛先アドレスを認識し、あらかじめ設定されたアドレスを宛先アドレスとするフレームを取り込んで接続される制御プロセッサへ送る。また制御プロセッサ114～117から受け取ったデータ及び制御情報をフレーム形式のデータにしてループ133へ送出する。ループ133は、HIFC103～106、FC151及び制御プロセッサ114～117を端末としてFibre Channel Arbitrated Loop (FCAL) と呼ばれるトポロジカルなループ伝送路を形成する。ファイバチャネルの通信プロトコルについては、例えばANSI公開マニュアル「FIBRE CHANNEL PHYSICAL AND SIGNALLING (FC-PH) REV.4.3」に記載されている。

#### 【0020】

例えばPBC108は、HIFC103を介してホストコンピュータ100と接続され、制御プロセッサ114, 115及びPBC111と接続可能である。従ってホストコンピュータ100から送られた入出力要求のコマンドは、PBC108を介して制御プロセッサ114, 115又はPBC111を介して制御プロセッサ116, 117で処理することができる。同様にホストコンピュータ101から送られた入出力要求のコマンドは、PBC109を介して制御プロセッ

サ 114, 115 又は PBC 110 を介して制御プロセッサ 116, 117 で処理することができる。

#### 【0021】

本実施形態では、ファイバチャネル 141, ループ 125 及び 126 にも、ファイバチャネルインタフェースを採用している。従って図示していないが、制御プロセッサ 119, 120 とファイバチャネル 141 との間に FC 151 が介在する。

#### 【0022】

図 3 は、FCAL 管理情報 113 のデータ構成を示す図である。FCAL 管理情報 113 は、制御プロセッサ 114 ~ 117 の各々についてループ 133 を介して取り込むべきフレームと入出力処理の対象とするデバイス番号の範囲を設定するテーブルである。FCAL 管理情報 113 の各エントリは、制御プロセッサ 201、AL-PA (Arbitrated Loop Physical Address) 202 及び LUN (Logical Unit Number) 203 を有する。制御プロセッサ 201 は、制御プロセッサ 114 ~ 117 のいずれかの制御プロセッサの識別子である。AL-PA 202 は、その制御プロセッサに割り当てられたループ 133 上のアドレスである。LUN 203 は、制御プロセッサが入出力処理を分担する論理デバイス番号又は論理デバイス番号の範囲である。FCAL 管理情報 113 の情報は、サービスプロセッサ 131 からの指示により設定又は設定解除することができる。

#### 【0023】

図 4 は、制御プロセッサ 114 ~ 117 の処理の流れを示すフローチャートである。制御プロセッサ 114 ~ 117 は、各々周期的に FCAL 管理情報 113 から自プロセッサのエントリを読み込み、接続されている FC 151 に自プロセッサの AL-PA を設定し、変更があれば AL-PA を再設定する。FC 151 は、HIFC 103 及びループ 133 を介してホストコンピュータ 100 から送られるフレーム上の AL-PA を読み取り(ステップ 301)、接続されている制御プロセッサの AL-PA でなければ(ステップ 302 NO)、処理を終了する。接続されている制御プロセッサの AL-PA であれば(ステップ 302 YES)、その制御プロセッサに通知する。通知を受けた制御プロセッサ 114 ~ 117 の

いずれかは、FC 1 5 1 を介してフレームを読み込み(ステップ 3 0 3)、フレーム中の入出力コマンドの L U N が L U N 2 0 3 の範囲にあるか否かをチェックする(ステップ 3 0 4)。もし指定された L U N が L U N 2 0 3 の範囲になればホストコンピュータ 1 0 0 へエラー応答を返す。次に制御プロセッサは受け取った入出力コマンドに従って入出力要求を実行する(ステップ 3 0 5)。

#### 【 0 0 2 4 】

制御プロセッサ 1 1 4 ~ 1 1 7 は、入出力要求が書き込み要求であれば、ホストコンピュータ 1 0 0 からデータを受領し、キャッシュメモリ 1 2 2 上の該当するキャッシュスロットにデータを書き込み、書き込み要求処理を終了する。データを書き込むべきキャッシュスロットのスロット番号はデータに付された L B A (Logical Block Address) から計算される。またスロット番号に対応するキャッシュメモリ 1 2 2 上のメモリアドレスは、共通制御メモリ 1 1 2 上のキャッシュスロット管理テーブルから得られる。入出力要求が読み出し要求であれば、キャッシュメモリ 1 2 2 上に要求されたデータがある場合、キャッシュメモリ 1 2 2 から読み出し、ループ 1 3 3 及び H I F C 1 0 3 を介してホストコンピュータ 1 0 0 へ送る。キャッシュスロット管理テーブルを参照することによって目的のデータの有無を判定することができる。キャッシュメモリ 1 2 2 上に要求されたデータがなければ、共通制御メモリ 1 1 2 上のプロセッサ間連絡領域に読み出し要求を書き込み、目的のデータがキャッシュメモリ 1 2 2 上に格納されたと判定したとき、キャッシュメモリ 1 2 2 から読み出してホストコンピュータ 1 0 0 へ送る。

#### 【 0 0 2 5 】

制御プロセッサ 1 1 9, 1 2 0 は、キャッシュメモリ 1 2 2 を探索し、キャッシュスロットにドライブ 1 2 7 ~ 1 3 0 に書き込むべきデータがあれば、ファイバチャネル 1 4 1、D I F C 1 2 3, 1 2 4 及びループ 1 2 5, 1 2 6 を介してそのデータをドライブ 1 2 7 ~ 1 3 0 に書き込む。この書き込みは、ホストコンピュータ 1 0 0 と制御プロセッサ 1 1 4 ~ 1 1 7 との間の入出力要求の処理動作とは非同期に行われる。制御プロセッサ 1 1 9, 1 2 0 は、指定された L U N 及び L B A を物理デバイス番号と物理アドレスに変換して書き込み対象とするドラ

イブとドライブ内アドレスを決定する。制御プロセッサ 119, 120 は、共通制御メモリ 112 上のプロセッサ間連絡領域を参照し、データ読み出し要求があれば、該当するドライブ 127~130 からデータを読み出し、キャッシュメモリ 122 上の該当するキャッシュスロットに書き込み、キャッシュスロット管理テーブル上で該当するデータ有に更新する。

#### 【0026】

ドライブ 127~130 への入出力要求は、制御プロセッサ 119, 120 のいずれでも処理可能である。例えば制御プロセッサ 119 またはファイインタフェースのループ 125 が故障等で使用できないとき、制御プロセッサ 120 とファイインタフェースのループ 126 によって入出力要求を処理できる。これにより障害発生時もドライブ 127~130 への入出力を停止することなく入出力要求を実行できる。

#### 【0027】

制御プロセッサ 114, 115, 116 及び 117 は、お互いの制御プロセッサの状態を監視している。各制御プロセッサは、一定周期で共通制御メモリ 112 に現在時刻を書き込む。それを他の制御プロセッサが一定周期で参照し、前回参照時と時刻の差分があるか否かをチェックし、差分がなければ当該制御プロセッサは停止していると判断する。停止状態を見つけた制御プロセッサは、FCAL 管理情報 113 からその制御プロセッサの管理情報を引き継ぎ、処理を継続する。例えば制御プロセッサ 114 が制御プロセッサ 115 の停止を検出したとする。このとき制御プロセッサ 114 は、FCAL 管理情報 113 を図 5 のように書き換える。これにより制御プロセッサ 114 は、制御プロセッサ 115 が処理していた LUN 10~19 に関する入出力要求を引き継いで処理することが可能である。

#### 【0028】

また制御プロセッサ 114~117 は、処理した入出力要求の処理数を計数し、一定周期で共通制御メモリ 112 に格納する。他の制御プロセッサは、この処理数を参照し、処理数の多い制御プロセッサを見つけ出し、より処理数の少ない制御プロセッサの処理数を上げることにより、各制御プロセッサの処理数を平均



化する。例えば制御プロセッサ 117 が制御プロセッサ 116 の処理数の低下と制御プロセッサ 115 の処理数の上昇を検出したとする。制御プロセッサ 117 は、F C A L 管理情報 113 を図 6 のように書き換える。ただしループ 133 を介して A L - P A が E 8 を有するフレームが制御プロセッサ 116 に伝送されるように、P B C 108 ~ 111 のうちの該当する P B C のスイッチ制御を変更しなければならない。これにより制御プロセッサ 116 は、L U N 10 ~ 19 及び L U N 20 ~ 29 に関する入出力要求を処理することが可能となり、制御プロセッサ間の処理数を平均化し、負荷分散を実現できる。

#### 【0029】

なお制御プロセッサの管理する L U N 203 のうち一部の L U N のみを他の制御プロセッサが引き継いで処理を継続することも可能である。例えば制御プロセッサ 115 の管理する L U N 10 ~ 19 のうち制御プロセッサ 116 が L U N 15 ~ 19 のみを引き継ぐことも可能である。このとき F C A L 管理情報 113 は図 7 のように書き換えられる。ただし L U N 15 ~ 19 について A L - P A 202 と L U N 203 との対応づけが変更となるので、制御プロセッサは、ホストコンピュータ 100, 101, 102 にこの変更を通知しなければならない。

#### 【0030】

なお上記制御プロセッサ 114 ~ 117 の処理の流れは、ファイバチャネルインタフェースによってディスク制御装置 107 と接続されるホストコンピュータ 100 に関する入出力要求の処理について説明した。ホストコンピュータ 101, 102 は、ファイバチャネルインタフェースとは異なるインタフェースによってディスク制御装置 107 と接続するので、H I F C 104 および 105 は、ホストコンピュータ 101 及び 102 から受け取った入出力コマンドをファイバチャネルインタフェースに従うフレームの形式に変換した後に、ループ 133 を介して制御プロセッサ 114 ~ 117 へ送信する。これによってホストコンピュータ 101 および 102 から送られる入出力要求の処理は上記処理と同様となる。H I F C 104 は、E S C O N (Enterprise System Connection) と呼ばれるインタフェースに従うコマンド、制御情報、データとファイバチャネルインタフェースに従うコマンド、制御情報、データとの間の変換をする機能を有する。また H

I F C 1 0 5 は、S C S I に従うコマンド、制御情報、データとファイバチャネルインタフェースに従うコマンド、制御情報、データとの間の変換をする機能を有する。このようにホストインタフェースの変換機能を有する H I F C をディスク制御装置 1 0 7 に搭載することによって、ホストインタフェースの種類に関係なく、任意のホストコンピュータをディスク制御装置 1 0 7 に接続することができる。

#### 【 0 0 3 1 】

また本実施形態はドライブ 1 2 7 ~ 1 3 0 がディスクドライブの場合について説明したが、D I F C 1 2 3, 1 2 4 を変更することによって磁気テープ装置やフロッピディスク・ドライブを接続することができる。また D I F C に、S C S I とファイバチャネルインタフェースの変換機能を設けることによって、ループ 1 2 5, 1 2 6 を S C S I によるケーブルに置き換えることもできる。

#### 【 0 0 3 2 】

本実施形態のディスク制御装置 1 0 7 によれば、ホストコンピュータ 1 0 0 から送られる入出力要求は制御プロセッサ 1 1 4 ~ 1 1 7 のうちのどの制御プロセッサでも処理できる。従ってホストコンピュータ 1 0 0 と H I F C 1 0 3 との間およびループ 1 3 3 のデータ転送速度に応じてホストコンピュータ 1 0 0 からの入出力要求の数が多い場合には、制御プロセッサ 1 1 4 ~ 1 1 7 すべてがホストコンピュータ 1 0 0 からの入出力要求を処理することが可能であり、制御プロセッサの台数がより少ない場合に比べてスループットが向上する。同様にホストコンピュータ 1 0 1 及び 1 0 2 から送られる入出力要求は、各々制御プロセッサ 1 1 4 ~ 1 1 7 のうちのどの制御プロセッサでも処理できる。このようにホストコンピュータ 1 0 0, 1 0 1 及び 1 0 2 がループ 1 3 3 及び制御プロセッサ 1 1 4 ~ 1 1 7 を共有するので、従来のようにホストコンピュータ 1 0 0, 1 0 1 及び 1 0 2 ごとに共通バスに接続するホストインタフェース部が独立している構成に比べて機構の分割損をなくし、記憶制御装置の性能向上を図るとともにコスト／性能比の向上を図ることができる。

## 【 0 0 3 3 】

## 【発明の効果】

以上述べたように本発明によれば、ホストコンピュータから送られる入出力要求を複数の制御プロセッサによって並列処理するとともに、制御プロセッサ間で負荷の分散をするので、記憶制御装置の性能向上を図ることができる。特に高速のファイバチャネルの性能を充分生かすことができる。また 1 つの制御プロセッサが障害により停止した場合に、他の制御プロセッサが障害制御プロセッサの処理を引き継ぐので、信頼性の高い記憶制御装置を提供できる。

## 【 0 0 3 4 】

さらに本発明の記憶制御装置は、複数種類のインタフェースをもつホストコンピュータを接続可能であるとともに、複数のホストコンピュータが記憶制御装置内部のファイバチャネルループ及び制御プロセッサを共有するので、コスト／性能比のよい記憶制御装置を提供できる。また異なる種類の記憶媒体の駆動装置を接続可能な記憶制御装置を提供できる。

## 【図面の簡単な説明】

## 【図 1】

実施形態の記憶サブシステムの構成図である。

## 【図 2】

実施形態のループ 1 3 3 と関連機構の構成図である。

## 【図 3】

実施形態の F C A L 管理情報 1 1 3 のデータ構成を示す図である。

## 【図 4】

実施形態の制御プロセッサ 1 1 4 ～ 1 1 7 の処理の流れを示すフローチャートである。

## 【図 5】

制御プロセッサが停止したときに書き換えた F C A L 管理情報 1 1 3 の例を示す図である。

## 【図 6】

制御プロセッサの負荷の不平衡を検出したときに書き換えた F C A L 管理情報

1 1 3 の例を示す図である。

【図 7】

制御プロセッサの負荷の不均衡を検出したときに書き換えた他の F C A L 管理情報 1 1 3 の例を示す図である。

【図 8】

従来の記憶制御装置の構成例を示す図である。

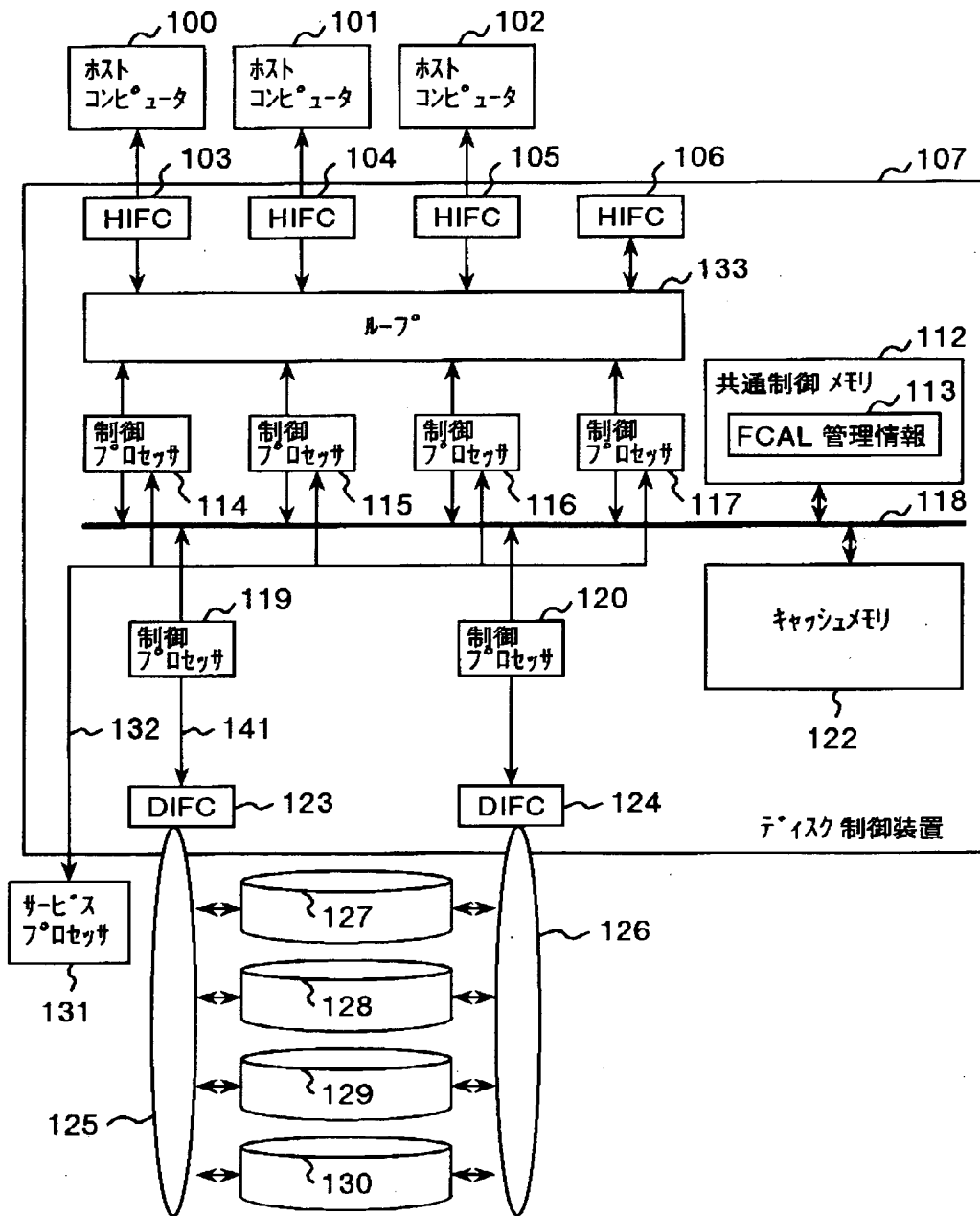
【符号の説明】

1 0 0, 1 0 1, 1 0 2 : ホストコンピュータ、1 0 3, 1 0 4, 1 0 5, 1 0 6 : ホストインタフェースコントローラ、1 0 7 : ディスク制御装置、1 1 3 : F C A L 管理情報、1 1 4 ~ 1 1 7, 1 1 9 ~ 1 2 0 : 制御プロセッサ、1 2 3, 1 2 4 : ドライブインタフェースコントローラ、1 2 7, 1 2 8, 1 2 9, 1 3 0 : ドライブ

【書類名】 図面

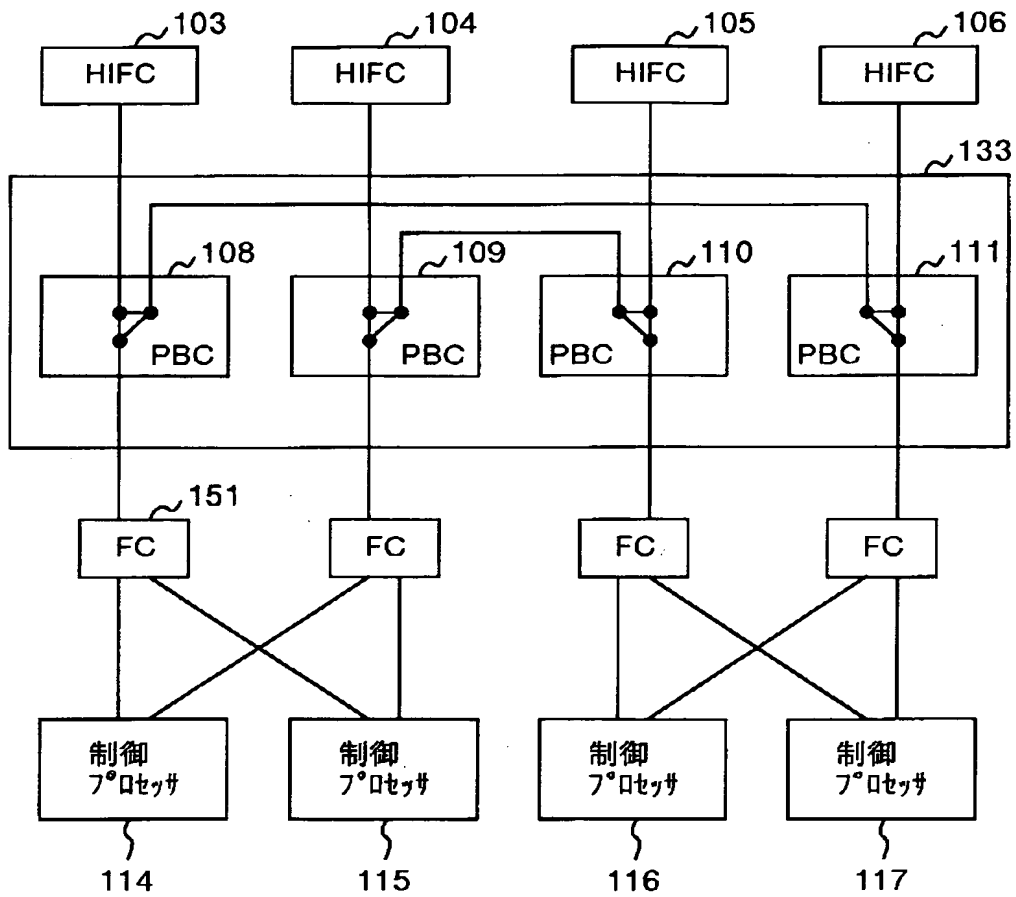
【図 1】

図 1



【図 2】

図 2



【図 3】

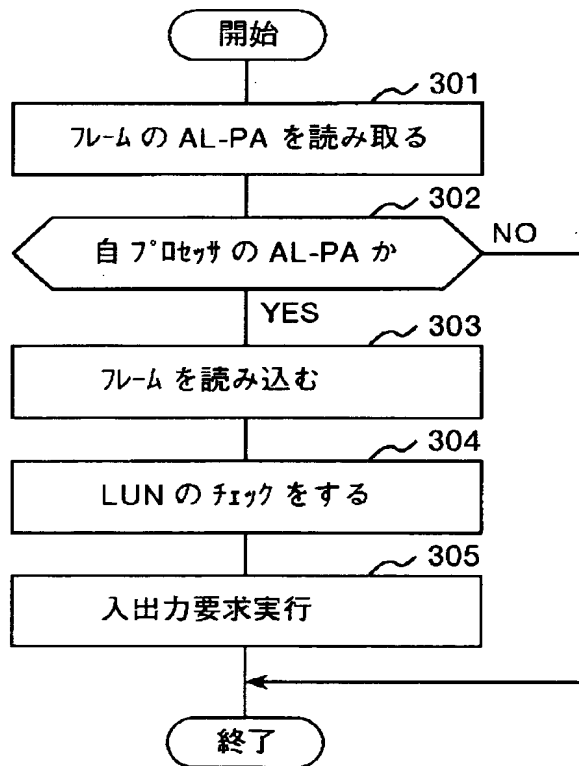
図 3

1 1 3 : F C A L 管理情報

制御プロセッサ番号	AL-PA	LUN
1 1 4	E F	0 - 9
1 1 5	E 8	1 0 - 1 9
1 1 6	E 4	2 0 - 2 9
1 1 7	E 2	3 0 - 3 9

【図 4】

図 4



【図 5】

図 5

113 : FCAL 管理情報

制御プロセッサ番号	AL-PA	LUN
114	EF	0-9
114	E8	10-19
115	-	-
116	E4	20-29
117	E2	30-39

【図 6】

図 6

113 : FCAL 管理情報

制御プロセッサ番号	AL-PA	LUN
114	EF	0-9
115	-	-
116	E8	10-19
116	E4	20-29
117	E2	30-39

【図 7】

図 7

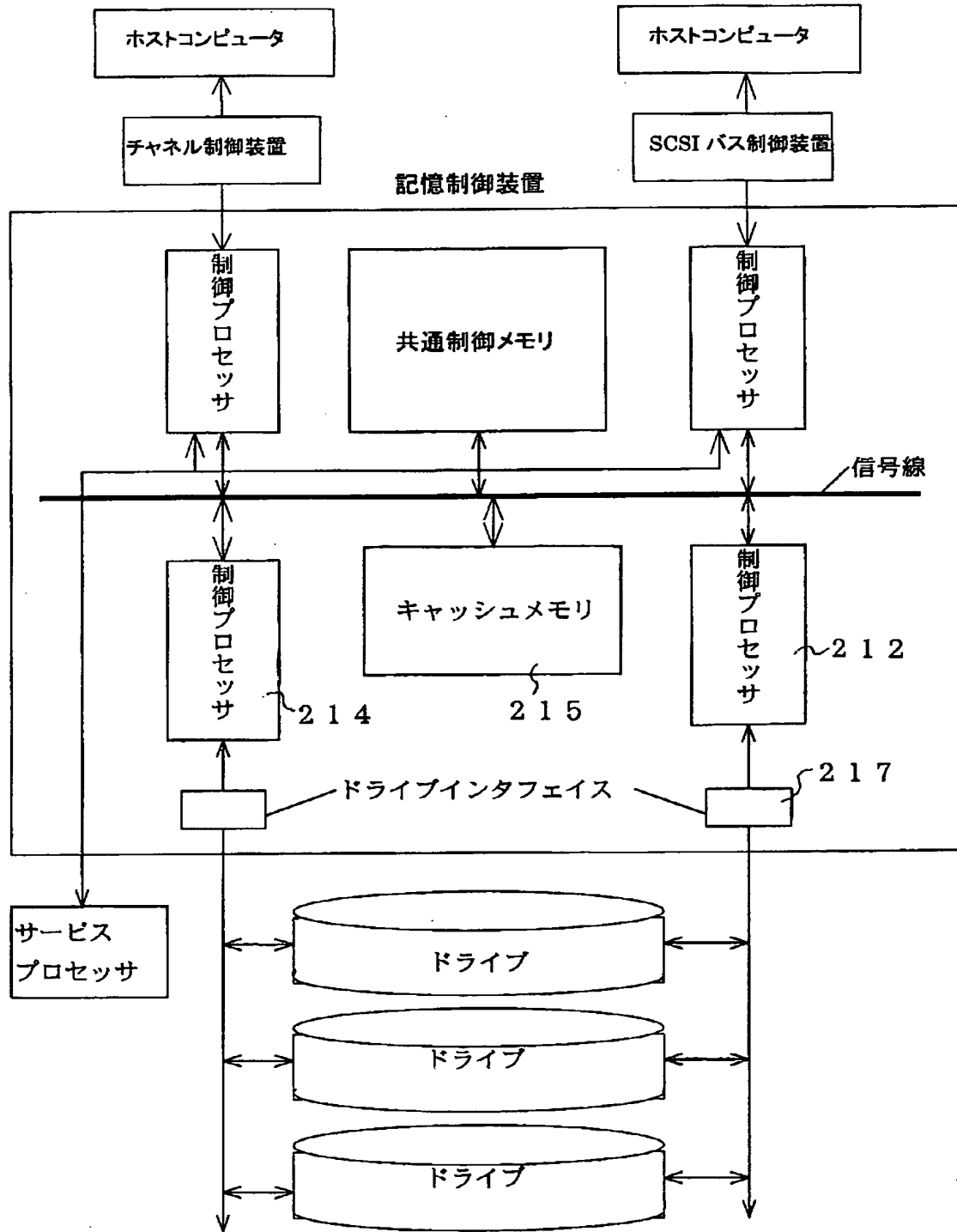
113 : FCAL 管理情報

制御プロセッサ番号	AL-PA	LUN
114	EF	0-9
115	E8	10-14
116	E4	15-29
117	E2	30-39



【図 8】

図 8



【書類名】 要約書

【要約】

【課題】 記憶制御装置の性能向上、特にファイバチャネルのもつ高速データ転送を生かすような高性能をもつとともに、信頼性の向上を図る。また複数種類のインタフェースをもつホストコンピュータを接続可能とする。

【解決手段】 ループ133はファイバチャネルインタフェースをもつ共通のループ伝送路である。HIFC103, 104, 105は、各々インタフェースの異なるホストコンピュータ100, 101, 102と接続し、必要に応じファイバチャネルインタフェースとの間の変換機能を有する。制御プロセッサ114～117は、HIFC103～105によって共有されるプロセッサである。制御プロセッサ114～117は、各々FCAL管理情報113を参照し、ループ133を流れるフレームのうち設定されたアドレスをもつフレームを取り込み、設定されたLUN範囲の入出力要求の処理をする。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台4丁目6番地
氏 名	株式会社日立製作所

For: The Patent Application

Our case No. NT0083US

**THE PRIOR ART REFERENCES CITED IN THE SPECIFICATION**

1. Japanese Unexamined Patent Publication No. Hei 05-143242



Home



Search



List

☐ Include**MicroPatent® PatSearch FullText: Record 1 of 1**

Search scope: US; EP; WO; JP

Years: 1976-2000 Patent/Publication No.: JP05143242

[no drawing available]

[Download This Patent](#)[Family Lookup](#)[Go to first matching text](#)**JP05143242  
MAGNETIC DISK SYSTEM  
HITACHI LTD**

Inventor(s): ARASAWA NOBUYUKI ; TOYODA MITSURU ; OYAMA MITSUO

Application No. 03308829, Filed 19911125, Published 19930611

**Abstract:** PURPOSE: To constitute the system so that a host computer can execute an access to the magnetic disk system by setting an optimal function to every channel.

**CONSTITUTION:** Between a disk controller 2 and a disk device group 12, a cache circuit 14 and a data retrieving circuit 15, a switching means for switching arbitrarily a connecting relation of outputs of the cache circuit 14 and the data retrieving circuit 15, and a disk controller side bus 5, and a switching means for switching arbitrarily inputs of the cache circuit 14 and the data retrieving circuit 15, and a data bus 13 of the disk device group are provided.

COPYRIGHT: (C)1993,JPO&amp;Japio

Int'l Class: G06F00306; G06F01314

MicroPatent Reference Number: 000196526

COPYRIGHT: (C)JPO



Home



Search



List

☐ Include

For further information, please contact:

[Technical Support](#) | [Billing](#) | [Sales](#) | [General Information](#)

(19)日本国特許庁 (J P) (12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平5-143242

(43)公開日 平成5年(1993)6月11日

(51)IntCl <sup>4</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	3 0 1 C	7165-5B		
		L 7165-5B		
13/14	3 1 0 F	7230-5B		

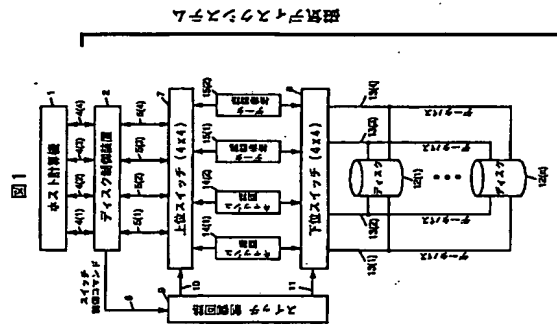
審査請求 未請求 請求項の数4(全16頁)

(21)出願番号	特願平3-308229	(71)出願人	000005108 株式会社日立製作所
(22)出願日	平成3年(1991)11月25日	(72)発明者	荒澤 伸幸 東京都千代田区神田駿河台四丁目6番地
		(72)発明者	豊田 漢 株式会社日立製作所中央研究所内 東京都国分寺市東恋ヶ窪1丁目280番地
		(72)発明者	大山 光男 株式会社日立製作所中央研究所内 東京都国分寺市東恋ヶ窪1丁目280番地
		(74)代理人	弁理士 小川 勝男 株式会社日立製作所中央研究所内 東京都国分寺市東恋ヶ窪1丁目280番地

(54)【発明の名称】 磁気ディスクシステム

(57)【要約】

【構成】 ディスク制御装置2と、ディスク装置群1.2との間に、キヤッシュ回路1.4とデータ検索回路1.5、キヤッシュ回路1.4とデータ検索回路1.5の出力とディスク制御装置2との接続関係で任意に切り換えられる切り換え手段と、キヤッシュ回路1.4とデータ検索回路1.5の入力とディスク装置群1.4とデータ検索回路1.5の間に任意に切り換えられる切り換え手段とを設ける。  
【効果】 ホスト計算機は、チャネルごとに最適な性能を設定して磁気ディスクシステムにアクセスできる。



【特許請求の範囲】

【請求項1】 複数のチャネルを持つホスト計算機の制御下において、複数のディスク制御装置を制御するディスク制御装置と、前記ディスク制御装置の制御下にあり、前記複数のチャネルを持つディスク装置群とからなる磁気ディスクシステムにおいて、前記ディスク制御装置と前記ディスク装置群との間に、前記ディスク装置群からのデータ読み出しを高速化する高速読み出し処理手段と、データの検索を行うデータ検索処理手段と、前記二つの処理手段の出力と前記ディスク制御装置のデータバスの接続関係を任意に切り換えることのできる、ディスク制御装置の切り換え手段と、前記二つの処理手段の出力と前記ディスク制御装置のデータバスの接続関係を任意に切り換えることのできるディスク装置群の切り換え手段とを設けたことを特徴とする磁気ディスクシステム。

【請求項2】 複数のチャネルを持つホスト計算機の制御下において、複数のディスク制御装置を制御するディスク制御装置と、前記ディスク制御装置の制御下にあり、前記複数のチャネルを持つディスク装置群とからなる磁気ディスクシステムにおいて、前記ディスク制御装置と前記ディスク装置群との間に、前記ディスク装置群からのデータ読み出しを高速化する高速読み出し処理手段と、前記高速読み出し手段と同一機能をもつ予備の高速読み出し処理手段と、前記二つの処理手段の出力と前記ディスク制御装置のデータバスの接続関係を任意に切り換えることのできるディスク制御装置の切り換え手段と、前記二つの処理手段の出力と前記ディスク装置群のデータバスの接続関係を任意に切り換えることのできるディスク制御装置の切り換え手段とを設けたことを特徴とする磁気ディスクシステム。

【請求項3】 複数のチャネルを持つホスト計算機の制御下において、複数のディスク制御装置を制御するディスク制御装置と、前記ディスク制御装置の制御下にあり、前記複数のチャネルを持つディスク装置群とからなる磁気ディスクシステムにおいて、前記ディスク制御装置と前記ディスク装置群との間に、データの検索を行う複数のディスク装置群と、前記データ検索処理手段の出力と前記ディスク制御装置のデータバスの接続関係を任意に切り換えることのできるディスク制御装置の切り換え手段と、前記ディスク装置群の任意のデータバスを複数回、前記データ検索処理手段の出力に接続することのできるディスク装置群の切り換え手段とを設けたことを特徴とする磁気ディスクシステム。

【請求項4】 複数のチャネルを持つホスト計算機の制御下において、複数のディスク制御装置を制御するディスク制御装置と、前記ディスク制御装置の制御下にあり、前記複数のチャネルを持つディスク装置群とからなる磁気ディスクシステムにおいて、前記ディスク制御装置と前記ディスク装置群との間に、データの

の検索を行なう複数のデータ検索処理手段と、前記データ検索処理手段の出力と前記ディスク制御装置のデータバスの接続関係を任意に切り換えることのできる前記ディスク制御装置の切り換え手段と、前記データ検索処理手段の入力とディスク装置群のデータバスの接続関係を任意に切り換えることのできるディスク装置群の切り換え手段とを設けたことを特徴とする磁気ディスクシステム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、複数のチャネルに接続される磁気ディスクシステムに係り、特に、各チャネルからのアクセスにおいて、それぞれ異なる性能、あるいは、機能を選択できる磁気ディスクシステムに関する。

【0002】

【従来の技術】 磁気ディスクシステムの通込分野が広がることにより、一台の磁気ディスクシステムに対し、入出力性能の高速化・多様化が要求されてきている。例えば、ホスト計算機上で実行される複数のジョブが、それぞれチャネルを介して、一台の磁気ディスクシステムにアクセスする場合、各ジョブごとにアクセスパターンが異なり、磁気ディスクシステムに要求する好ましい性能、機能（例えば高速スループット、高速データ転送、あるいは検索機能など）が異なることも多い。これに対し、従来は、例えば特開昭63-223822号公報に開示されているように、単一の性能を向上させるものであった。

【0003】 図9は、従来の技術の説明図である。図9において、ディスク制御装置2とディスク制御装置105の間に、MPU101とインタフェース回路102とが設けられ、MPU101はインタフェース回路102を介して、ディスク制御装置2とディスク制御装置105の間の並列処理アダプタ100を設け、ホスト計算機1からの命令を、MPU101で判断し、バッファ切り換え回路を制御する事によって、ディスクからデータバッファ、データバッファからディスク制御装置へのデータ転送の並列処理を実現し、データ転送の高速化を図っている。

【0004】

【発明が解決しようとする課題】 図10は従来技術の効果と問題点を説明するフローチャートである。例えば、ホスト計算機1から、データのリード/ライト命令が、図9の並列処理アダプタ100に与えられたとき（ステップ110）、命令の処理が始まり（ステップ111）、データのリード/ライト命令なので（ステップ112）、並列処理アダプタによりスループットが向上する（ステップ113）。しかし、磁気ディスクシステムに求められている他の機能（例えばデータ検索など）については、性能の向上を図ることができない（ステップ114）。

【0005】このように従来の技術は、特定の機能を磁気ディスクシステムに付加する技術には有効であるが、一台のディスクシステムに複数の性能、あるいは機能、例えば、キャッシュを用いた高速読み出し機能とデータ検査の高速化機能の二つの機能を持たせ、かつ必要に応じて使い分けることができる磁気ディスクシステムについては考慮されていなかった。このため、このような場合、性能、機能の異なる複数の磁気ディスクシステムを使用する必要があった。

【0006】また従来技術では、性能の向上のための手段、例えば、図9に示すようなデータバッファが、ディスク装置群のバスに固定して設けられているので、バスがメモリが故障した場合、ディスク装置群のデータがアクセス不能となる場合があり、コンピュータシステムに大きな影響を与える。

【0007】また、検査するデータ量が増加しているため、データ検査時間に占める磁気ディスクシステムからホスト計算機への、データ転送時間の割合が大きくなってきており、データ転送時間の短縮が望まれている。

【0008】本発明の目的は、一台のディスクシステムに複数の性能、性能を持たせ、必要に応じて柔軟に使い分けができるような磁気ディスクシステムを実現することにある。

【0009】本発明の他の目的は、性能向上の手段、新しい機能の付加手段に際しては、性能向上の場合にも、障害発生を回避することにより、障害がシステムに重大な影響を及ぼさない磁気ディスクシステムを実現することにある。

【0010】【課題を解決するための手段】上記の目的を達成するために、複数のチャネルを介してホスト計算機に接続されるディスク制御装置と、複数のデータバスを持つデータ検査装置群からなる磁気ディスクシステムにおいて、前記データ検査装置群と前記ディスク装置群との間に、前記ディスク装置群からのデータ読み出しを高速化させるためのキャッシュと、読み出したデータを検査し、必要ならデータのみをホスト計算機に送るための検査手段と、前記キャッシュと、前記データ検査手段の出力を前記ディスク制御装置と、任意のチャネルに接続させるための上位バスと、前記キャッシュと前記データ検査手段の出力を、前記ディスク装置群の任意のバスと接続するための下位バスと、二つのスイッチを前記ディスク制御装置からの指示により制御するスイッチ制御回路を設け、ホスト計算機で実行されるジョブが必要とする機能を選択できるようにした。

【0011】さらに、キャッシュを、少なくとも一個の代替のキャッシュを含めて構成し、前記キャッシュに障害が発生した場合は、障害が発生したキャッシュを代替のキャッシュに切り換えるようにした。

【0012】さらに、複数のデータ検査手段を設け、か

つ下位バスに、磁気ディスク装置群の、一つのデータバスのデータを複数のデータ検査手段に出力する機能と、複数のデータバスに同一データを出力する機能とを設けることにより、前記ディスク装置群から読み出したデータを複数のデータ検査手段に同時に出力して検査能力を高めるようにした。

【0013】さらに、複数のデータバスを持つディスク装置に、複数のデータバスに同一データを出力する手段と、複数のデータ検査手段を設け、検査するデータを複数の前記データ検査手段に同時に出力して検査能力を高めるようにした。

【0014】

【作用】本発明では、図12に示すように、ディスク制御装置が、スイッチ制御回路に切り換えタイミングを与えることにより、上位バスと下位バスを制御して、データ検査の場合には、データ検査回路を選択し、高速データ転送が必要な場合には、キャッシュ回路を選択して使用できる。すなわち、ホスト計算機が必要とするアクセス特性に合わせて、一台の磁気ディスクシステムを、異なる複数の性能、機能を有する磁気ディスクシステムとして使用できるため、効率の良いデータアクセスを行なうことができる。また、キャッシュ回路とデータ検査回路は複数のチャネルに対応して並列に動作可能であり、あたかも、異なる機能を持つ複数のディスクシステムのように動作可能となる。

【0015】また、キャッシュ回路とデータ検査回路を、キャッシュ回路と代替キャッシュ回路とで構成することにより、キャッシュ回路に障害が生じたときは代替のキャッシュ回路に切り換えて、障害によるディスクシステム性能低下、あるいは停止を防ぐことができるため、ディスクシステム性能の向上が図れる。

【0016】また、複数のデータ検査手段に、ディスク装置群から読み出したデータを同時に出力することにより、それぞれのデータ検査手段で別々のデータ検査を行って行なうことができるため、データ検査能力が向上する。

【0017】

【実施例】本発明の実施例について、図面を参照して説明する。図1は本発明の第一の実施例を示す図、図2はキャッシュ回路14の構成図である。図3はデータ検査回路15の構成を示す図である。

【0018】図1において、1は磁気ディスクシステムに接続されるホスト計算機、2は磁気ディスクシステム全体を制御するディスク制御装置、4(1)～4(4)は、ホスト計算機1とディスク制御装置2間で、コマンドや、データの送受信を行なうために用いるチャネル、5(1)～5(4)はディスク制御装置2とキャッシュ回路14、データ検査回路15との間でコマンドや、データの送受信を行なうために用いるバス、14(1)～14(2)はディスク装置12(1)～12(n)から読み出したデータをキャッシュに記憶することにより、

たところに特徴がある。

【0022】ホスト計算機1にディスク装置12(1)からデータを読み込む場合、ホスト計算機1は、チャネル4(1)～4(4)のうち、空きのチャネルを用いて、ディスク制御装置2にデータ読み込みを行うロードコマンドを発行する。ディスク制御装置2はバス5(1)～5(4)、キャッシュ回路14(1)～14(2)、データバス13(1)～13(4)を調べ、この時、例えばバス5(1)とキャッシュ回路14(1)とデータバス13(1)が空いていた場合、ディスク制御装置2は、スイッチ制御回路9にスイッチ制御コマンド6を発行して、上位スイッチ7と下位スイッチ8を切り換えて、バス5(1)とキャッシュ回路14(1)とデータバス13(1)を接続する。

【0023】次に、キャッシュ回路14(1)に制御コマンドを発行して、データを読み込んだ時の動作を設定する。この時、キャッシュ回路14(1)ではコマンド6、バス20に制御コマンドを取り込み、コマンドデコード21で解釈を行い、キャッシュ制御回路22にキャッシュに読み込むデータ値を設定する。この後、ディスク制御装置2は一つのディスク装置12(1)にロードコマンドを発行する。ディスク装置12(1)は、シーケンス動作と回転待ちを行いデータを読み出す。この時、ディスク装置12(1)は、キャッシュ回路14(1)に設定したデータ値を読み出す。そしてディスク制御装置2に指定されたデータバス13(1)を用いてキャッシュ回路14(1)にデータを転送する。

【0024】キャッシュ回路14(1)内では入力バッファ22にデータを一時取り込み、キャッシュ制御回路23によってデータをキャッシュメモリ24に書き込む。更にデータを出力バッファ25に書き込み、バス(1)を通してディスク制御装置2に転送する。ディスク制御装置2は、このデータを送ってホスト計算機1にデータを転送する。更に、このデータを使ってホスト計算機1にデータを取り込まれているキャッシュ回路14(1)にデータの回転待ちをナシめ、最初の読み出しに比べて、ディスクの回転待ちナシめ、時間分早く読むことができる。このためキャッシュ回路14(1)～14(2)は連続してデータを読むときに効果が大きい。

【0025】更に、磁気ディスクシステムで読み出したデータに対して、データ検査を行ない、必要なデータのみをホスト計算機1に転送する場合、ホスト計算機1はチャネル4(1)～4(4)のうち、空きのチャネルを用いて、ディスク制御装置2に検査コマンドと検査データを送る。ディスク制御装置2はバス5(1)～5(4)、データ検査回路15(1)～15(2)、データバス13(1)～13(4)を調べる。この時、例えばバス5(1)とデータ検査回路15(1)とデータバス13(1)が空いていた場合、ディスク制御装置2

キャッシュにヒットした場合のデータ読み込み時間を早くする。データ検査回路15(1)～15(2)は読み出したデータの検査を行なうデータ検査回路、7はバス5(1)～5(4)をキャッシュ回路14(1)～14(2)の出力、およびデータ検査回路15(1)～15(2)の出力に任意に接続する上位スイッチ、8はキャッシュ回路14(1)～14(2)の入力、およびデータ検査回路15(1)～15(2)の入力とディスク装置12(1)～12(n)とを、データバス13(1)～13(4)を介して、任意に接続する下位スイッチ、9はディスク制御装置2から出力されるスイッチ制御コマンド6を解釈して、上位スイッチ7と下位スイッチ8における接続を制御するスイッチ制御回路である。

【0019】図2において、20はディスク制御装置2から発行されたコマンドを入力するコマンドバッファ、21はコマンドデコード、22はコマンドデコードで解釈された内容がキャッシュに関連したコマンドであったときにキャッシュを制御するキャッシュ制御回路、25はキャッシュから読み出したデータを保持する出力バッファ、23はディスク装置12(1)～12(2)から読み出したデータを保持する入力バッファ、24は入力バッファ23の内容をキャッシュ制御回路22の制御によって記憶するキャッシュメモリ、26は出力バッファ25のバリエーション(データ書き込み時にディクティチェック回により付加)ディクティチェックを行なうバリエーション回路、27は出力バッファにバリエーションが発生した時に出力される出力エラー信号である。

【0020】図3において、30はディスク制御装置2から発行されたコマンドを入力するコマンドバッファ、31はコマンドデコード、32はコマンドデコードで解釈された内容がデータ検査に関連したコマンドであったときに比較器35を制御する比較器制御回路、33はディスク装置12(1)～12(n)から読み出したデータを保持する入力バッファ、34はディスク制御装置2から送られてきた比較データと保持する比較データレジスタ36の内部の条件が一致したときにそのデータを保持するデータレジスタ37は出力バッファ36のバリエーション(データ書き込み時にディスク制御装置2に付加)ディクティチェックを行なうバリエーション回路、38は出力バッファにバリエーションが発生した時に出力される出力エラー信号である。

【0021】本実施例では、ホスト計算機1へのデータの読み込みと検査が、同時に発生するよう磁気ディスクシステムにおいて、キャッシュ回路14(1)～14(2)とデータ検査回路15(1)～15(2)を、上位スイッチ7と下位スイッチ8を用いて切り換えることにより高速処理を行う磁気ディスクシステムを実現し

は、スイッチ制御回路9にスイッチ制御コマンド6を発行して、上位スイッチ7と下位スイッチ8を切り換え、バス5(1)とデータ検査回路15(1)とデータバス13(1)を接続する。更にデータ検査回路15(1)に制御コマンドと検査データを転送して、データを読み込んだ時の動作を設定する。この時、データ検査回路15(1)ではコマンドバッファ30にコマンドを取り込み、コマンドデコーダ31で解除を行い、比較器32に動作を設定し、比較データレジスタ34に検査データを書き込む。

【0026】この後、ディスク制御装置2はディスク装置12(2)にリードコマンドを発行する。ディスク装置12(2)はデータの読み出しを行い、ディスク制御装置2から指定されたデータバス13(1)を用いてデータ検査回路15(1)にデータを転送する。データ検査回路15(1)内では入力バッファ33にデータを一時取り込み、比較データレジスタ34内のデータと比較器35によって比較を行ない、条件を満たしたデータだけを一致データレジスタ36に書き込み、さらにバス5(1)を通してディスク制御装置2に転送する。

【0027】ディスク制御装置2は、このデータをチャネル4(1)～4(4)のうち空きのチャネルを使ってホスト計算機1に送る。従来はデータ全てをホスト計算機1に送る。必要データだけを転送することにより、不要なデータを取り除いて、必要なデータだけを転送することができるため、転送するデータ量が少なく、データ転送時間が短縮される。

【0028】第一の実施例では、データの読み込みと検査の二つの異なる処理を、ホスト計算機1がキャッシュ回路14(1)～14(2)とデータ検査回路15(1)～15(2)を切り換えて使うことにより行っているが、上位スイッチ7、下位スイッチ8で切り換えて使用される機能実現回路は、磁気ディスクシステムが使用される機能実現回路に限定されるものではない。

【0029】図4は本発明の第二の実施例を示す図である。

【0030】図4において、14(1)～14(4)はディスク装置12(1)～12(n)から読み出したデータをキャッシュに記憶することにより、ヒットした場合のデータ読み込み時間を早くするキャッシュ回路、14(5)は、予備のキャッシュ回路、41はバス5(1)～5(4)とキャッシュ回路14(1)～14(5)の出力を任意に接続する上位スイッチ、42はキャッシュ回路14(1)～14(5)の入力とディスク装置12(1)～12(n)とを任意に接続する下位スイッチである。

【0031】本実施例では、ホスト計算機1へのデータの読み出しを行なう磁気ディスクシステムにおいて、キ

ャッシュ回路14(1)～14(4)に異常が発生した場合、予備のキャッシュ回路14(5)に切り換えることにより、性能低下、あるいは使用不能になることなく、処理を続けられる耐障害性の高い磁気ディスクシステムを実現する。

【0032】ホスト計算機1にディスク装置12(1)からデータを読み出す場合、ホスト計算機1はチャネル4(1)～4(4)のうち、空きのチャネルを用いて、ディスク制御装置1にデータ読み込みを行うリードコマンドを発行する。ディスク制御装置2はバス5(1)～5(4)、キャッシュ回路14(1)～14(4)、データバス13(1)～13(4)を調べ、この時、例えば、バス5(1)とキャッシュ回路14(1)とデータバス13(1)が空いていた場合、ディスク制御装置2は、スイッチ制御回路9にスイッチ制御コマンド6を発行して、上位スイッチ41と下位スイッチ42を切り換えて、バス5(1)とキャッシュ回路14(1)とデータバス13(1)を接続する。

【0033】次に、キャッシュ回路14(1)に制御コマンドを発行して、データを読み込んだ時の動作を設定する。この時、キャッシュ回路14(1)ではコマンドバッファ20に制御コマンドを取り込み、コマンドデコーダ21で解除を行い、キャッシュ制御回路22にキャッシュに読み込むデータ量を設定する。この後、ディスク制御装置2はディスク装置12(1)にリードコマンドを発行し、ディスク装置12(1)は、キャッシュ回路14(1)に、設定されたデータ量を読み出す。そして、データバス13(1)を用いて、キャッシュ回路14(1)にデータを転送する。

【0034】キャッシュ回路14(1)内では入力バッファ23にデータを一時取り込み、キャッシュ制御回路22によってデータをキャッシュメモリ24に書き込む。更にデータを入力バッファ25に書き込み、バス5(1)を介してディスク制御装置2に転送する。この時、出力バッファ25のバリディチェックを行なったバリディチェック回路26がバリディエラー27を検出した場合、キャッシュ回路14(1)内に何らかの異常が発生したことになり、キャッシュ回路14(1)は異常を検出をディスク制御装置2に報告する。ディスク制御装置2はスイッチ制御回路9にスイッチ制御コマンド6を発行して異常が発生したキャッシュ回路14(1)を切り換え、予備キャッシュ回路14(5)に切り換える。この後ディスク制御装置2は、再度ディスク装置12(1)にリードコマンドを発行して処理を続ける。ディスク制御装置2は、正しく読み込まれたデータをチャネル4(1)～4(4)のうち空きのチャネルを使ってホスト計算機1にデータを送る。

【0035】以上に説明した第2の実施例では、データの読み込み処理を、ホスト計算機1がキャッシュ回路14(1)～14(4)を使うことにより実現している。

もしキャッシュ回路14(1)～14(4)のうち、どれかが故障した場合、予備キャッシュ回路14(5)に切り換える事により、磁気ディスクシステムが性能低下、あるいは使用不能になることなく処理を続けることができる。このように、第二の実施例では、耐障害性の高い磁気ディスクシステムを実現している。

【0036】図5は本発明の第三の実施例を示す図である。図6は全バス出力スイッチ51の構成図である。

【0037】図5において、15(1)～15(4)は読み込んだデータの検査を行なうデータ検査回路、51はデータ検査回路15(1)～15(4)とディスク装置12(1)～12(n)とを、バス13(1)～13(4)を介して任意に接続し、ディスク装置12(1)～12(n)が読み出したデータを複数のデータ検査回路に入力する全バス出力スイッチである。

【0038】本実施例では、磁気ディスクシステムにおいて、読み出したデータに対して、複数のデータ検査回路15(1)～15(4)に同一データを入力してデータ検査処理を並列に実行する磁気ディスクシステムを実現することにより特徴がある。本実施例の磁気ディスクシステムにおいて、同一データについて三種のデータ検査を行なった結果を、ホスト計算機1に読み込む場合、ホスト計算機1は、チャネル4(1)～4(4)のうち空きのチャネルを用いて、ディスク制御装置2に検査コマンドと三種の異なる検査データを転送する。ディスク制御装置2はバス5(1)～5(4)、データバス13(1)～13(4)を調べ、この時、例えば、バス5(1)と三つのデータ検査回路15(1)～15(3)とデータバス13(1)が空いていた場合、ディスク制御装置2は、スイッチ制御回路9にスイッチ制御コマンド6を発行して、上位スイッチ7と全バス出力スイッチ51を切り換えて、バス5(1)と三つのデータ検査回路15(1)～15(3)とデータバス13(1)とを接続する。

【0043】本実施例では、磁気ディスクシステムにおいて、読み出したデータに対して、データ検査を行ない、必要なデータのみをホスト計算機1に転送する場合、ディスク装置12(2)から読み出したデータを全バスデータ検査回路71(1)～71(2)を用いて、複数のデータ検査回路15(1)～15(4)に出力することができる全バスディスク装置である。

【0044】本実施例の磁気ディスクシステムにおいて、同一データについて三種のデータ検査を行なったデータ検査処理を並列に行う磁気ディスクシステムを実現することにより特徴がある。

【0045】本実施例の磁気ディスクシステムにおいて、同一データについて三種のデータ検査を行なったデータ検査回路15(1)～15(4)のうち空きのチャネルを用いて、ディスク制御装置2に検査コマンドと三種の異なる検査データを転送する。この時、三つのデータ検査回路15(1)～15(3)ではコマンドバッファ30にコマンドを取り込み、コマンドデコーダ31で解除を行い、比較器制御回路32の動作を設定し、比較データレジスタ34に比較データを書き込む。この後、ディスク制御装置2はディスク装置12(2)に読み込みコマンドを発行し、ディスク装置12(2)は、データを読み出す。この時ディスク装置12(2)はディスク制御装置2に指定されたデータバス13(1)を用いてデータ検査回路15(1)～15(3)にデータを転送するが、全バス出力スイッチ51を通して

て三つのデータ検査回路15(1)～15(3)に同一データが転送される。

【0039】それぞれのデータ検査回路15(1)～15(3)内では入力バッファ33にデータを一時取り込み、比較データレジスタ34内に比較データと比較器35によって比較を行ない、条件を満たしたデータを一致データレジスタ36に書き込む。この後三つのデータ検査回路15(1)～15(3)内の一致データをディスク制御装置2に転送する。

【0040】ディスク制御装置2は、このデータをチャネル4(1)～4(4)のうち空きのチャネルを使ってホスト計算機1にデータを転送する。従来はデータをホスト計算機1に読み込んでから検査を行っていたが、データ検査回路15(1)～15(4)を設けることにより、不要なデータを取り除いて、必要なデータだけを転送することができるため、転送するデータ量が少なく、データ転送時間が短縮される。更に、本実施例では三つの検査を同時に行なうため、検査時間も短縮される。

【0041】図7は本発明の第四の実施例を示す図である。図8は図7における全バスディスク装置71(1)～71(n)の構成図である。

【0042】図7において、15(1)～15(4)は読み込んだデータの検査を行なうデータ検査回路、71(1)～71(2)は読み出したデータを任意のバス13(1)～13(4)に出力することができる全バスディスク装置である。

【0043】本実施例では、磁気ディスクシステムにおいて、読み出したデータに対して、データ検査を行ない、必要なデータのみをホスト計算機1に転送する場合、ディスク装置12(2)から読み出したデータを全バスデータ検査回路71(1)～71(2)を用いて、複数のデータ検査回路15(1)～15(4)に同一データ検査処理を並列に行う磁気ディスクシステムを実現することにより特徴がある。

【0044】本実施例の磁気ディスクシステムにおいて、同一データについて三種のデータ検査を行なったデータ検査回路15(1)～15(4)のうち空きのチャネルを用いて、ディスク制御装置2に検査コマンドと三種の異なる検査データを転送する。この時、三つのデータ検査回路15(1)～15(3)ではコマンドバッファ30にコマンドを取り込み、コマンドデコーダ31で解除を行い、比較器制御回路32の動作を設定し、比較データレジスタ34に比較データを書き込む。この後、ディスク制御装置2はディスク装置12(2)に読み込みコマンドを発行し、ディスク装置12(2)は、データを読み出す。この時ディスク装置12(2)はディスク制御装置2に指定されたデータバス13(1)を用いてデータ検査回路15(1)～15(3)にデータを転送するが、全バス出力スイッチ51を通して



る。更に三つのデータ検索回路15(1)～15(3)に制御コマンドと三つの異なる検索データを転送して、データを読み込んだ時の動作を設定する。この時、三つのデータ検索回路15(1)～15(3)ではコマンドバッファ30にコマンドを取り込み、コマンドデコーダ31で解読を行い、比較制御回路32の動作を設定し、比較データレジスタ34に比較データを書き込む。この後、ディスク制御装置2は全バスディスク装置7(1)にリードコマンドを発行する。全バスディスク装置7(1)は、データを読み出す。この時、全バスディスク装置7(1)はディスク制御装置2に指定されたデータバス13(1)～13(3)全てに同一データを出力して、データ検索回路15(1)～15(3)にデータを転送する。それぞれのデータ検索回路15(1)～15(3)内では入力バッファ33にデータを一時取り込み、比較データレジスタ34にある比較データと比較器35によって比較を行ない、条件を満たしたデータだけを一致データレジスタ36に書き込む。この後三つのデータ検索回路15(1)～15(3)内の一致データをディスク制御装置2に転送する。ディスク制御装置2は、このデータをチャネル4(1)～4(4)のうち空きのチャネルを使ってホスト計算機1にデータを転送する。従来はデータ全てをホスト計算機1に読み込んでから検索を行っていたが、データ検索回路15(1)～15(4)を設けることにより、不要なデータを取り除いて、必要なデータだけを送ることができ、転送するデータ量が少なく、データ転送時間が短縮される。更に、本実施例では3つの検索を同時にこなうため、検索時間も短縮される。

【0045】図11は、本発明の実施例の効果を説明するフローチャートである。ホスト計算機からある処理が発行された場合(ステップ121、122)、ディスク制御装置において処理の種類を調べる(ステップ123)。リードの場合は、更にキャッシュ回路に故障が無いかチェックし(ステップ124)、故障の無い場合はスイッチを切り換えてキャッシュ回路を選択し(ステップ126)、処理を実行する。この時、キャッシュ回路はリード性能を向上させる(ステップ127)。もし、故障があった場合は、予備のキャッシュ回路に切り換えて(ステップ128)、同様にリード性能を向上させる(ステップ129)。ディスク制御装置において処理の種類を調べた時(ステップ123)、データ検索の場合は、データ検索回路に故障が無いかどうかをチェックし(ステップ125)、無い場合はスイッチを切り換えてデータ検索回路を選択し(ステップ130)、処理を実行する。この時データ検索回路は検索性能を向上させる(ステップ131)。もし、故障があった場合は、予備

のデータ検索回路に切り換えて(ステップ132)、検索性能を向上させることができる(ステップ133)。

【0046】他の効果として、ディスク装置から読み出したデータを複数のデータ検索回路に入力してデータ検索処理を並列に行なうことにより、データ検索を高速に行う磁気ディスクシステムを実現することもできる。

【0047】

【発明の効果】本発明では、ホスト計算機へのデータの読み込みと検索が、頻繁に発生するような磁気ディスクシステムにおいて、キャッシュ回路とデータ検索回路をスイッチを用いて切り換え、キャッシュ回路を使用することにより、キャッシュにヒットした場合は、ディスクの回転待ち+シーク時間分早く読むことができるので、特にシーケンシャルリードの場合は特にヒットの確率が高く、高速にデータ転送を行なうことができる。データ検索の場合は、データ検索回路により、不要データをホスト計算機に送らないので、ホスト計算機へのデータ転送時間が短縮される。すなわち、本発明によれば、チャネルごとに、複数の異なる性能、あるいは機能から、必要となる性能、あるいは機能を選択して同時に使用できる。また、異なる機能を持つ複数のディスクシステムのように動作可能となり、磁気ディスクシステムにおいて、効率のよいアクセスを行なうことができ、また、上記の回路をキャッシュ回路と予備のキャッシュ回路で構成することにより、キャッシュ回路に障害が発生した場合にも、性能低下、あるいは使用不能になることなく処理を続けられ、耐障害性の高い磁気ディスクシステムを実現することができる。

【図面の簡単な説明】

【図1】本発明の第一の実施例を示すブロック図。

【図2】キャッシュ回路のブロック図。

【図3】データ検索回路のブロック図。

【図4】本発明の第二の実施例を示すブロック図。

【図5】本発明の第三の実施例を示すブロック図。

【図6】全バス出力スイッチのブロック図。

【図7】本発明の第四の実施例を示すブロック図。

【図8】全バスディスク装置のブロック図。

【図9】従来例の説明図。

【図10】従来例の効果と問題点を表すフローチャート。

【図11】本発明の効果を説明するフローチャート。

【図12】本発明の作用を説明するタイムチャート。

【符号の説明】

1…ホスト計算機、2…ディスク制御装置、7…上位スイッチ、8…下位スイッチ、9…スイッチ制御回路、5…全バス出力スイッチ、71(1)～71(n)…全バスディスク。

【図1】

図1

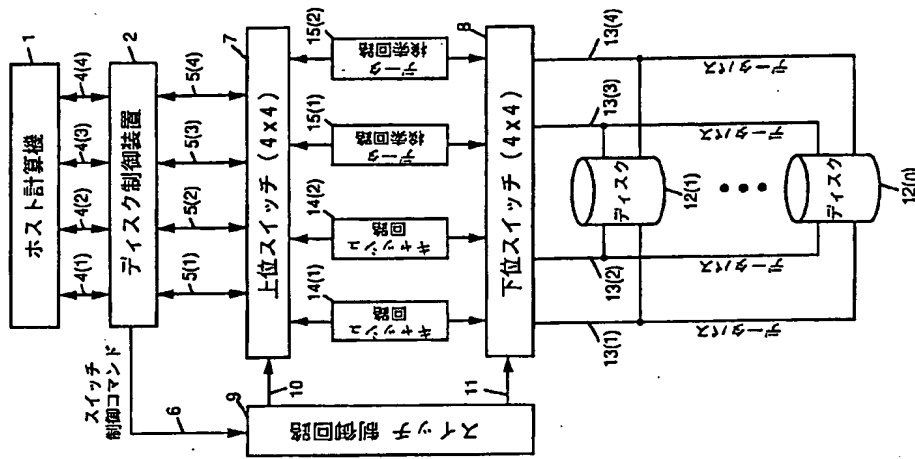
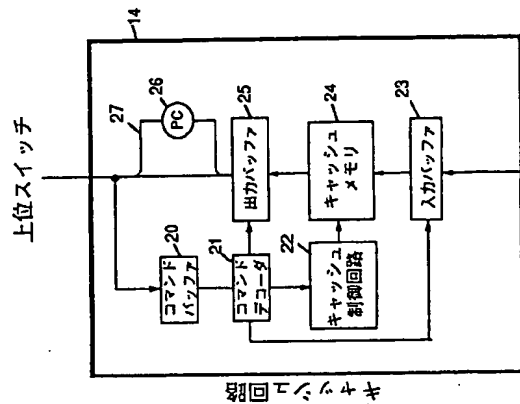
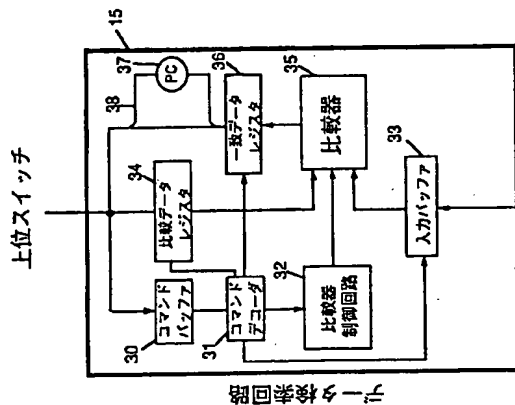


図1

【图2】

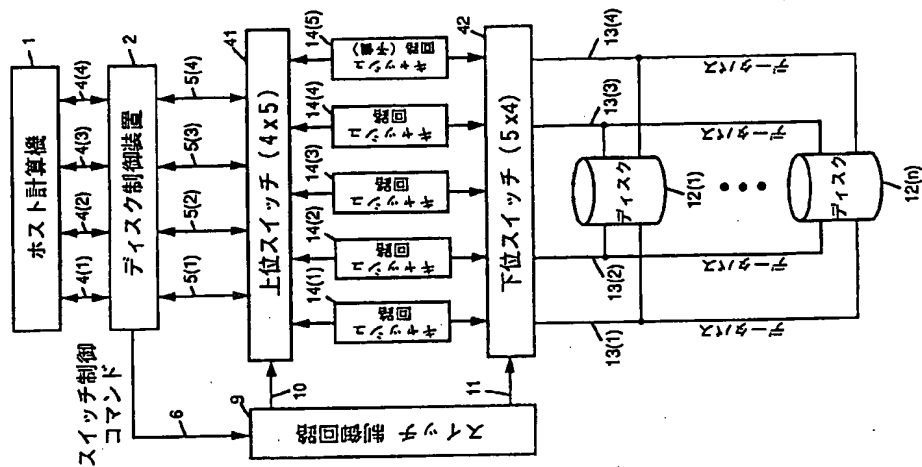


【图3】

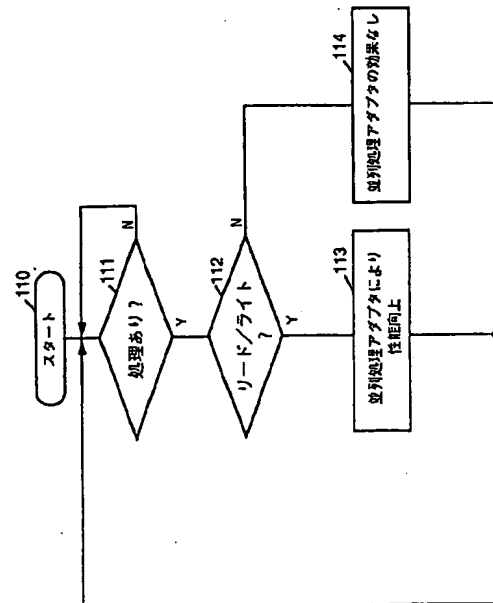


2

3  
✕

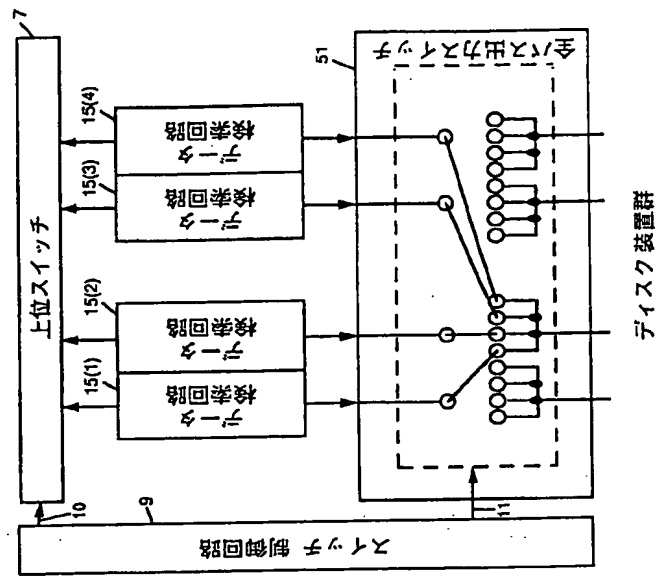
4  
✕

【図10】



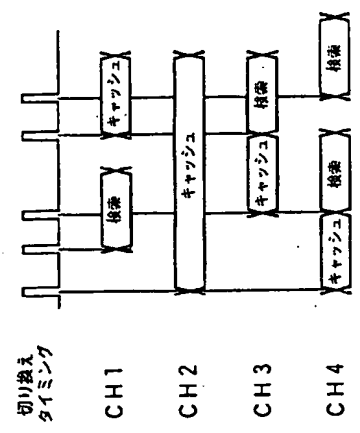
【図6】

図 6



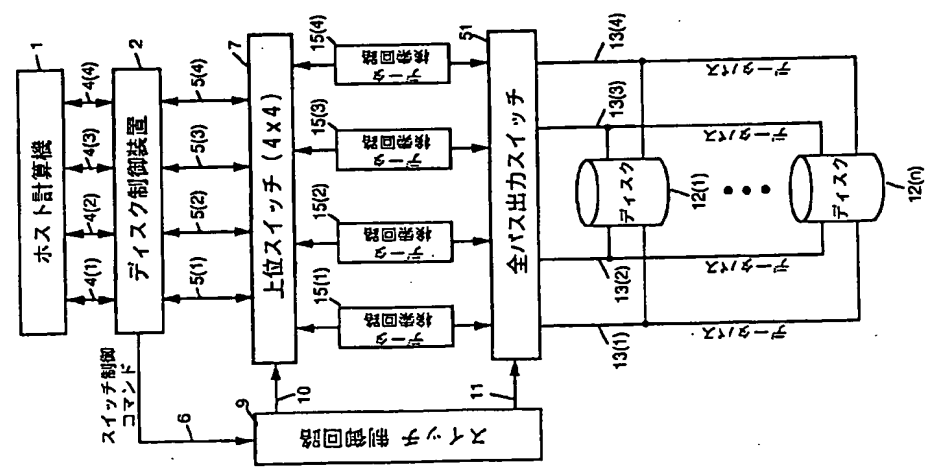
【図12】

図 12



【図5】

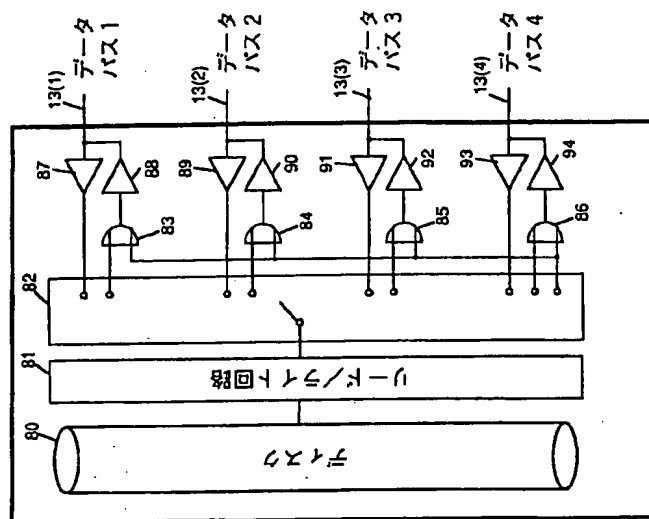
図 5



磁気ディスクシステム

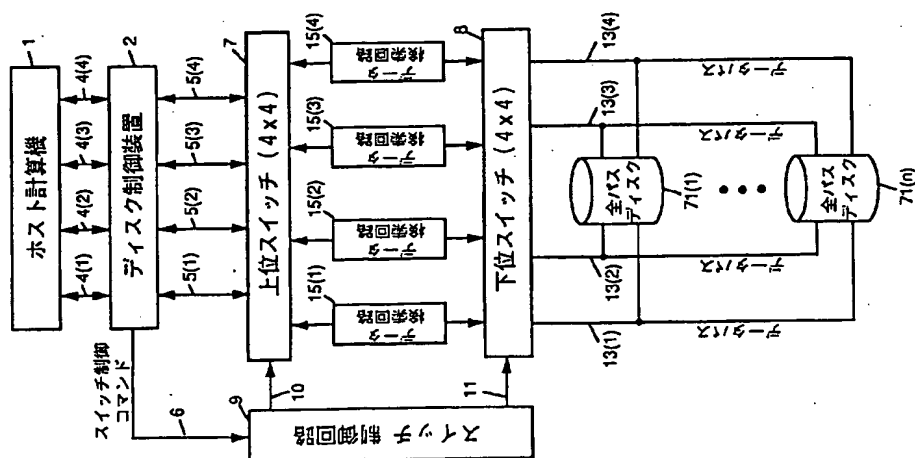
【図8】

図8



【図7】

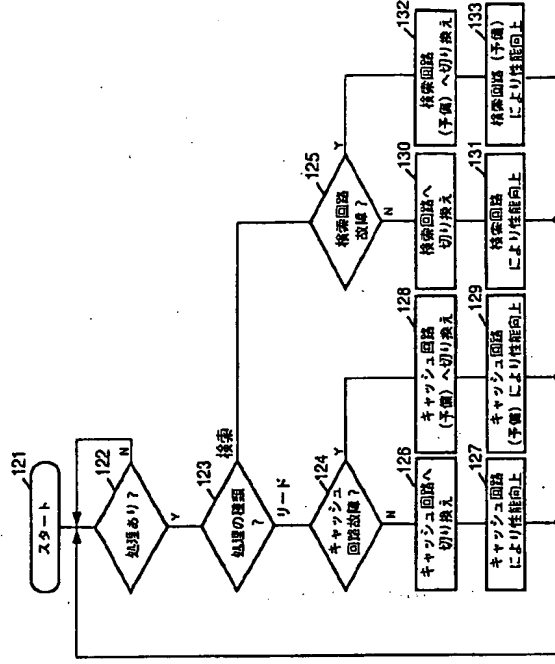
図7



磁気ディスクシステム

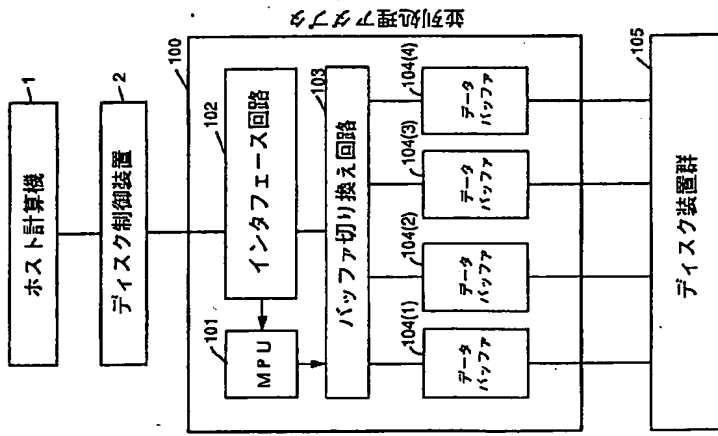
【図11】

図11



【図9】

図9



並列処理プログラマ